



Context matters: the probabilistic grammar of international varieties of English

Benedikt Szmrecsanyi and Melanie Röthlisberger



KU Leuven

Quantitative Lexicology and Variational Linguistics

Introduction



Cultural context matters

about

- the extent to which,
- the way in which,
- and the reasons because of which

syntactic variation patterns are **indigenized** in different cultures/regions using dialects of the same language.

⇒ “**lectalization**”



Background

“Exploring probabilistic grammar(s) in varieties of English around the world”

(<http://wwwling.arts.kuleuven.be/qlvl/ProbGrammarEnglish.html>)



Work program

1. explore 3 syntactic alternations across 9 varieties of English
2. use the variationist method (Labov 1982) to study probabilistic factors constraining the alternations
3. check for significant differences between varieties; confirm via rating-task experiments



Varieties of English covered



Data source

- International Corpus of English = ICE
- 1 mio words of text per variety
- 60% spoken, 40% written English
- 12 different registers



The English dative alternation

- (1)
- a. We sent [the president]_{recipient} [a letter]_{theme}
(the ditransitive dative)
 - b. We sent [a letter]_{theme} to [the president]_{recipient}
(the prepositional dative)



The English dative alternation

- (1)
- a. We sent [the president]_{recipient} [a letter]_{theme}
(the ditransitive dative)
 - b. We sent [a letter]_{theme} to [the president]_{recipient}
(the prepositional dative)

numerous known constraints: end-weight, animacy, pronominality, definiteness, info status, . . .



Indigenization



Classic indigenization effects in the World Englishes literature

“indigenization of language structure mostly occurs at the interface between **grammar** and **lexis**, affecting the syntactic behavior of certain lexical elements. Individual words, typically high frequency items, adopt characteristic but marked usage and complementation patterns.”

(Schneider 2007: 46; emphasis ours)



Classic indigenization effects in the World Englishes literature

“indigenization of language structure mostly occurs at the interface between **grammar** and **lexis**, affecting the syntactic behavior of certain lexical elements. Individual words, typically high frequency items, adopt characteristic but marked usage and complementation patterns.”

(Schneider 2007: 46; emphasis ours)

(2) *no worries*



Classic indigenization effects in the World Englishes literature

“indigenization of language structure mostly occurs at the interface between **grammar** and **lexis**, affecting the syntactic behavior of certain lexical elements. Individual words, typically high frequency items, adopt characteristic but marked usage and complementation patterns.”

(Schneider 2007: 46; emphasis ours)

(2) *no worries*

(3) *this hair-style is called **as** ‘duck tail’*



“Probabilistic indigenization”

“the process whereby **stochastic patterns** of internal linguistic variation are **reshaped** by shifting usage frequencies in speakers of post-colonial varieties. To the extent that patterns of variation in a new variety A [...] can be shown to differ from those of the mother variety, we can say that the new pattern represents a novel, if gradient, development in the grammar of A.”

(Szmrecsanyi, Grafmiller, Heller, and Röthlisberger 2016: 133)



“Cognitive indigenization”

Probabilistic indigenization

+

“stress the **outcome** of probabilistic indigenization, namely the ‘**lectalization**’ or creation of distinct lects (that speakers are aware of) at the level of very subtle gradience.”

(Röthlisberger, Grafmiller, and Szmrecsanyi submitted:4; emphasis ours)



Indigenization effects in the English dative alternation



Today: the English dative alternation

- (4)
- a. We sent [the president]_{recipient} [a letter]_{theme}
(the ditransitive dative)
 - b. We sent [a letter]_{theme} to [the president]_{recipient}
(the prepositional dative)



Procedure

- retrieval of $N = 8,549$ interchangeable dative tokens from the corpus database
(hand-coding)
- rich annotation for > 10 language-internal constraints and language-external factors
(partially hand-coded)
- mixed-effects logistic regression analysis to identify probabilistic contrasts



Probabilistic contrast 1: recipient pronominality

recipient pronominality has a greater effect in Indian and Canadian English and a weaker effect in Jamaican English:

- (5) a. *send him a telegram* <ICE-IND:s1a-058>
 b. *send an email to you* <ICE-JA:w1b-007>



Probabilistic contrast 2: end weight

the effect of end-weight (short before long) is weaker in IndE and stronger in JamE (compared to all other varieties):

- (6) a. *causing a great deal of inconvenience to commuters*
<ICE-IND:s2b-014>
- b. *pay Michael Jordan millions of dollars*
<ICE-JA:s2b-032>



Probabilistic contrast 3: theme concreteness

concrete themes favor the prepositional dative more strongly in Canadian English than in other varieties:

- (7) a. *given the key to Helen* <ICE-CAN:S1A-058>
 b. *given Heidi my T-shirt* <ICE-GB:S1B-066>



Discussion



Two patterns

- the “core” probabilistic grammar underlying the dative alternation is largely stable
- differences wrt the degree of sensitivity that speakers of different varieties demonstrate towards some of the factors that constrain dative choice



Biases in speech production & planning

MacDonald (2013):

- **Easy First**: speaker select early those linguistic units that are easier to retrieve from long-term memory – “easier”: frequent, shorter, less syntactically complex, conceptually entrenched, and given in the discourse
- **Plan Reuse**: speakers tend to reuse previously heard syntactic plans and closely related structures that they retrieve from long-term memory



Whence the probabilistic differences?

differences might arise as the result of modifications in the cue strength of predictors through the constant reinforcement of structural patterns by Plan Reuse:



Whence the probabilistic differences?

differences might arise as the result of modifications in the cue strength of predictors through the constant reinforcement of structural patterns by Plan Reuse:

1. language and dialect contact
2. SLA and substrate influences
3. constructional and/or semantic changes



Language and dialect contact

- new combinations of syntactic constructions with lexical items (Schneider 2007)
- when new lexical items occur frequently enough in syntactic constructions, the abstractions of regularities that speakers make lead to changes in the constraints governing language structure
- constraints are learned during language acquisition (Ellis 2002: 144) and become part of speakers' grammatical knowledge (Gahl and Garnsey 2004)

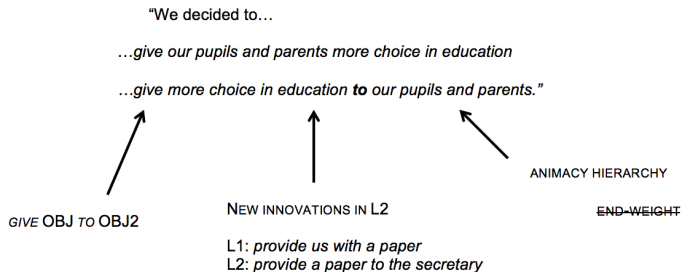


SLA and substrate influences

- in L2 varieties, increased usage of the more transparent syntactic variant – prepositional dative (Leufkens 2013; Siegel et al. 2014)
- changes in the strength of specific cues as variants are used by L2 speakers in contexts where L1 speakers would not use them
- transfer of cue strength (effect size of constraints) from L1 can lead to gradient shifts in linguistic preferences and changes in speakers' probabilistic L2 grammar (MacWhinney 1997: 29)



SLA and substrate influences



SLA and substrate influences

Can you...

...pass me the keys, please.

...pass the keys to me, please.

Bill should have...

...sent Marc the letter.

...sent the letter to Marc.

"We decided to...

...give our pupils and parents more choice in education

*...give more choice in education **to** our pupils and parents."*

GIVE OBJ TO OBJ2

NEW INNOVATIONS IN L2

L1: *provide us with a paper*

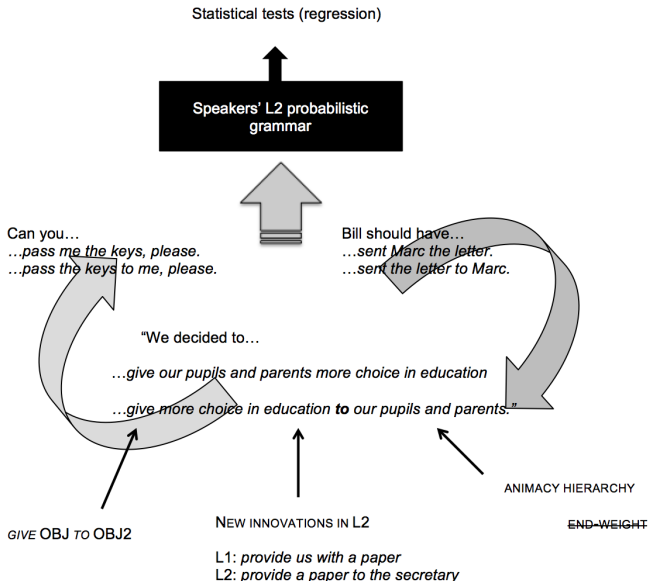
L2: *provide a paper to the secretary*

ANIMACY HIERARCHY

END-WEIGHT



SLA and substrate influences



Constructional and/or semantic changes

- speakers use ditransitive/prepositional datives in different ways in different contexts ➡ reflected in the lexical items that fill the syntactic slots
- range of lexical items might be more diverse in one variant as compared to the other, and this difference in diversity might differ from variety to variety (Dodson and Tomasello 1998; McDonough 2006)
- the constant reinforcement of these diverging usage patterns through Plan Reuse can result in diverging statistical regularities





Conclusion



Building bridges

1. language is used in cultural and/or regional context
2. cultural/regional differences engender probabilistic indigenization effects through entrenchment
3. entrenched indigenization leads to increasing awareness and lectalization
(= formation of distinct lects)



Lectalization in three-dimensional space

pipeline:

- 1 regression model / variety
- ⇔ 9×20 coefficient matrix
- ⇔ Euclidean distance matrix
- ⇔ cluster analysis (Ward)
- ⇔ MDS



Thank you!

`benszm@kuleuven.be`

`melanie.rothlisberger@kuleuven.be`

`http://wwwling.arts.kuleuven.be/
qlvl/ProbGrammarEnglish.html`

This presentation is based upon work supported by an
Odysseus grant of the Research Foundation Flanders (FWO)
(grant no. G.0C59.13N).

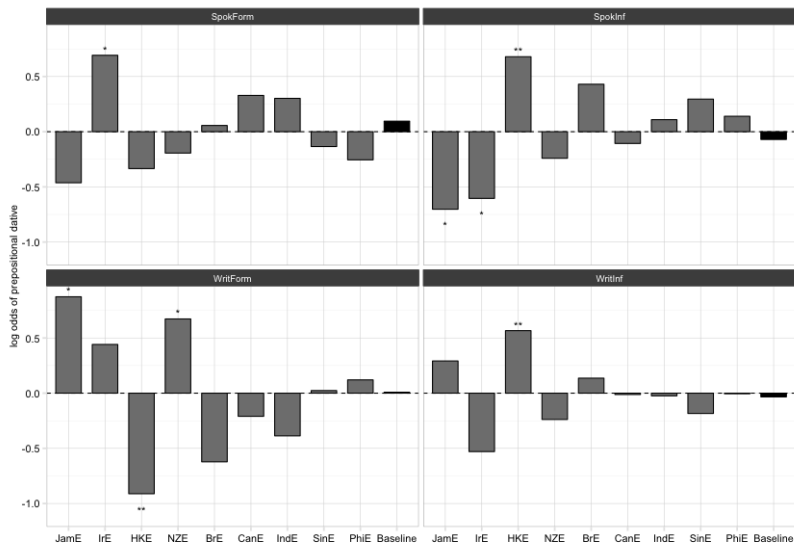


A word on register

- register differences ⇔ cultural norms
- indeed, plenty of register differences



A word on register



References I

- Dodson, K. and M. Tomasello (1998). Acquiring the transitive construction in English: the role of animacy and pronouns. *Journal of Child Language* 25(3), 605–622.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24, 143–188.
- Gahl, S. and S. Garnsey (2004). Knowledge of Grammar, Knowledge of Usage: Syntactic Probabilities Affect Pronunciation Variation. *Language* 80, 748–775.
- Geeraerts, D. (2010). Recontextualizing grammar: Underlying trends in thirty years of Cognitive Linguistics. In E. Tabakowska, M. Choinski, and L. Wiraszka (Eds.), *Cognitive Linguistics in Action: From Theory to Application and Back*, pp. 71–102. Berlin/New York: De Gruyter Mouton.



References II

- Labov, W. (1982). Building on empirical foundations. In W. Lehmann and Y. Malkiel (Eds.), *Perspectives on Historical Linguistics*, pp. 17–92. Amsterdam, Philadelphia: Benjamins.
- Leufkens, S. (2013). The transparency of creoles. *Journal of Pidgin and Creole Languages* 28(2), 323–362.
- MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology* 4, 1–16.
- MacWhinney, B. (1997). Second Language Acquisition and the Competition Model. In A. M. De Groot and J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*, pp. 113–142. Mahawa, NJ: Lawrence Erlbaum Associates.
- McDonough, K. (2006). Interaction and syntactic priming: English L2 speakers' production of dative constructions. *Studies in Second Language Acquisition* 28(2), 179–207.
- Röthlisberger, M., J. Grafmiller, and B. Szmrecsanyi. Cognitive indigenization effects in the English dative alternation.



References III

- Schneider, E. (2007). *Postcolonial English: Varieties Around the World*. Cambridge: Cambridge University Press.
- Siegel, J., B. Szmrecsanyi, and B. Kortmann (2014, January). Measuring analyticity and syntheticity in creoles. *Journal of Pidgin and Creole Languages* 29(1), 49–85.
- Szmrecsanyi, B., J. Grafmiller, B. Heller, and M. Röthlisberger (2016). Around the world in three alternations: modeling syntactic variation in varieties of English. *English World-Wide* 37(2).